DOCUMENT RESUME

ED 467 911 CE 083 554

AUTHOR	Kemple, James J.
TITLE	The Federal Role in Vocational and Technical Education at the Secondary Level: Principles for Moving toward a Greater Emphasis on Supporting Systematic Innovation and Rigorous Evaluation.
SPONS AGENCY	Office of Vocational and Adult Education (ED), Washington, DC.
PUB DATE	2002-02-00
NOTE	21p.
CONTRACT	ED-99-CO-0160
AVAILABLE FROM	For full text (MS Word): http://www.ed.gov/offices/OVAE/HS/kemple0408.doc.
PUB TYPE	Opinion Papers (120)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	Accountability; Cost Effectiveness; Educational Finance; *Educational Improvement; *Educational Innovation; Educational Policy; Educational Research; Federal Aid; *Federal Government; *Government Role; *Predictive Validity; Program Evaluation; Secondary Education; Technical Education; *Vocational Education
IDENTIFIERS	Department of Education; *No Child Left Behind Act 2001

ABSTRACT

This paper argues that the federal government's role in vocational and technical education (VTE) should include a greater emphasis on systematically testing promising strategies and interventions and subjecting them to rigorous evaluations of their effects, implementation, costs, and benefits. Section 1 presents a rationale for rethinking the federal role in VTE at the secondary level and highlights several potential benefits of moving to a greater focus on systematic innovation and rigorous evaluation. Section 2 outlines a set of guiding principles for what systematic innovation and rigorous evaluation ought to look like. Section 3 touches on several strategies that might be used to help shift the federal role toward a greater focus on these principles. (Throughout, the paper implies that a federal commitment to systematic innovation and rigorous evaluation should not be confined to VTE. It contends that the current administration's focus on ensuring that "no child is left behind" appears to be particularly well-suited to using federal resources to leverage state and local investments in promising strategies to improve schools and to rigorously evaluate them in ways that assess the impacts of the reforms and help develop accountability systems for further improvement.) (YLB)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 467 911

The Federal Role in Vocational and Technical Education at the Secondary Level: Principles for Moving Toward a Greater Emphasis on Supporting Systematic Innovation and Rigorous Evaluation

James J. Kemple

Manpower Demonstration Research Corporation

February 2002

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

This paper was prepared for the Office of Vocational and Adult Education, U.S. Department of Education pursuant to contract no. ED-99-CO-0160. The findings and opinions expressed in this paper do not necessarily reflect the position or policies of the U.S. Department of Education.

LE003554

BEST COPY AVAILABLE

The Federal Role in Vocational and Technical Education at the Secondary Level: Principles for Moving Toward a Greater Emphasis on Supporting Systematic Innovation and Rigorous Evaluation

James J. Kemple

Introduction

This paper argues that the federal government's role in vocational and technical education should include a greater emphasis on systematically testing promising strategies and interventions and subjecting them to rigorous evaluations of their effects, implementation, costs, and benefits. The objectives of this role shift would be to concentrate federal resources on problems and solutions that are not likely to be addressed by state and local education authorities alone and to fill important gaps in the nation's education policy-making capacity. By serving as the central purveyor of reliable evidence about what works and what does not, how effective programs work, and at what cost, the federal government is likely to exert greater leverage in shaping policy around excellence in vocational and technical education.

The first section of the paper presents a rationale for rethinking the federal role in vocational and technical education at the secondary level and highlights several potential benefits of moving to a greater focus on systematic innovation and rigorous evaluation. The second section outlines a set of guiding principles for what systematic innovation and rigorous evaluation ought to look like. The third section of the paper touches on several strategies that might be used to help shift the federal role toward a greater focus on these principles.

Throughout, the paper implies that a federal commitment to systematic innovation and rigorous evaluation should not be confined to vocational and technical education. The current administration's focus on ensuring that "no child is left behind" appears to be particularly well-suited to using federal resources to leverage state and local investments in promising strategies to improve schools and to rigorously evaluate them in ways that assess the impacts of the reforms and help develop accountability systems for further improvement.

Finally, the paper was prepared with the knowledge that some of the specific strategies and proposals it puts forth are likely to be constrained by current legislation and administrative regulations. The paper's purpose is to offer a rationale for a shift in the federal role and outline the beginnings of some strategies for moving in the direction of making greater investments in systematic innovation and rigorous evaluation. A

central challenge in pursuing such a shift, therefore, rests in utilizing the flexibility in current legislation and regulations and in helping to guide changes in Perkins and the federal role as they are brought up for debate and reauthorization in 2003.

Why focus the federal role on systematic innovation and rigorous evaluation?

Several prevailing issues come to mind as a rationale for suggesting that the federal government's role in education should be more focused on systematically testing promising innovative ideas and submitting them to rigorous evaluation.

- **The wide disbursement of the relatively small federal investment in vocational and technical education probably limits federal leverage to affect policies aimed at educational excellence and constrains its capacity to provide guidance to the field.**

The federal government provides only a small portion of the overall funding for elementary and secondary education in the United States (averaging between 7 and 9 percent annually). Federal investments in elementary and secondary schools, while often targeted to particular categories of students (special needs, low income, and so on) or programs (vocational/technical, math, science and technology, and so on), are typically blended unrecognizably with other local and state funding streams. This includes funding for vocational and technical education, which constitutes the largest federal investment in secondary education. In 1999, for example, allocations for vocational technical education from the Carl D. Perkins Vocational and Technical Education Act of 1998 (Perkins III) provided $750 million for secondary school students, more than any other federal source including Title I and TRIO combined. This is still less than 10 percent of all funding for vocational technical education at the secondary level. Further, most of this funding (approximately 85 percent) is driven by formulas that allocate federal resources based primarily on the number and characteristics of students covered by state and local education authorities.

Thus, by design, most of the federal investment in vocational and technical education is so widely dispersed that that is nearly impossible to trace its connection to key policy decisions or to assess whether it adds value to the quality of services being offered. As currently constructed, federal investments, outside the more limited research and demonstration allocations, are poorly positioned to inform or influence policy aimed at excellence in vocational and technical education. Shifting the federal role to better inform policy and practice would first require changing legislation so that funding criteria reflected priorities for addressing particular problems with promising solutions rather than population characteristics. Without such a change, research and demonstration funding should be viewed as particularly precious resources and should be applied with care.

- **Education policy and practice are not well informed by high-quality, coherent evaluation research, and these gaps are unlikely to be filled by state and local education authorities.**

Even though education is central to the civic, political, and economic life of the country, it has not been well informed by high-quality, crosscutting evaluation research compared, for example, with business and industry, health, national security, and even income security.[1] Certainly, variation in methodology and standards of evidence, have weakened the link between education research, policy, and practice. Much of the gap, however, may derive from the fact that education research is nearly as decentralized as the education system itself. In addition, data collection and assessment at the state and local level is typically aimed at accounting and assigning accountability for the use of public funds for education. Local school districts and even states do not typically have the expertise or the resources to invest in research that would be aimed at improving policies and practice more generally. The federal government can probably perform such a function best.

A large share of the federal investment in learning about vocational and technical education goes to the National Assessment of Vocational Education (NAVE) and the National Research and Dissemination Centers for Career and Technical Education (NRCCTE and NDCCTE). NAVE provides a useful scan of the landscape of vocational and technical education. It offers insight into the number and characteristics of these programs and provides documentation of the experiences and outcomes of students, teachers, and other stakeholders in vocational and technical education. NAVE, however, has not been well equipped to assess the impact of vocational and technical programs nor even of the federal investment in these programs. It has also had limited capacity to document the implementation of particular vocational and technical education strategies and was not set up to assess the relationship between net costs and benefits to students, schools, and the public.

NRCCTE and its predecessor the National Center for Research in Vocational Education (NCRVE) have served as useful clearinghouses for secondary data analyses and reviews of literature related to vocational and technical education. Both NRCCTE and NCRVE have also undertaken studies of particular versions or components of vocational and technical education such as curriculum integration approaches, school-to-work programs, work-based learning activities, and school-based enterprises. In many cases, however, these studies have been largely descriptive and were not designed to produce rigorous assessments of impacts. In cases where the studies did attempt to examine impacts on student or school outcomes, the research designs were not well equipped to resolve issues regarding internal validity or causal relationships between the interventions and changes in outcomes and did not provide evidence about the relationship between costs and benefits.

---

[1] See, for example, Lisa Towne, Richard J. Shevelson, and Michael J. Feur, editors, "Science, Evidence, and Inference in Education: Report of a Workshop" (Washington, DC: National Academy Press, 2001). Also, see Ellen Condliffe Lageman, *An Elusive Science: The Troubling History of Education Research* (Chicago: The University of Chicago Press, 2000) for a review of the historical context of the weak links between education research, policy, and practice.

Over the past several years, a number of expert panels have been convened to review the national capacity to supply teachers, administrators, and policymakers with reliable and accessible evidence about what works, and how it works, in education. Each has pointed out, to one degree or another, that individual states or districts, and even some individual research projects tend to undervalue the benefits or lessons to be derived from or offered to others. Without a federal role, dissemination of such lessons across levels of government is likely to languish, and the possibilities for a sustained, coherent, and crosscutting learning agenda would be greatly diminished. One such panel, organized by the National Research Council (NRC), outlined a set of principles for cultivating federal leadership in scientific inquiry and learning in education.[2] Given the national significance of such an effort and the likelihood that states and localities will not pick up the mantle, it seems natural to look to the federal government to play such a role. A fundamental concern, as noted in the NRC panel report, is that such an effort should be independent and largely insulated from political pressures that may compromise scientific integrity.

Finally, although the federal government has invested in efforts to test promising innovations and submit them to rigorous evaluation, such efforts have been rare and precious. The Office of Vocational and Adult Education's (OVAE) Tech Prep Demonstration Program (TPDP) and the Office of Educational Research and Improvement's (OERI) Comprehensive School Reform Demonstration (CSRD) program are two recent examples of federal support for the adaptation of promising practices that were to be coupled with evaluations of their implementation and impacts. These efforts are ongoing. If the core demonstration and evaluation activities of these initiatives can be executed with reasonably high fidelity to their intent, they may have the potential to guide a shift in the federal role toward larger investments in such efforts.

- **The current policy environment may provide a rare opportunity to refocus the federal role on systematic innovation and rigorous evaluations in vocational and technical education, as well as in others areas of education.**

Two related trends and events in the policy environment may be seen as coalescing in ways that would support a shift in the federal role in elementary and secondary education generally and vocational and technical education in particular. First, there has been a growing emphasis on the role that research should play in education policy at all levels. This has included a range of reviews of issues related to the standards of evidence that should be used to inform policy and of strategies to would promote "scientific" principles that would ensure higher quality education research. In some cases, the outcome of these reviews has taken the form of specific

[2] National Research Council, *Scientific Inquiry in Education* (Washington, DC: National Academy Press, 2001).

recommendations for research methods (as in H.R. 4875).[3] These recommendations tend to assume that high-quality research is equated with maximum use of highly acclaimed methodologies like random assignment. In other cases, however, the recommendations take the form of more general guiding principles for fostering a scientific culture among education researchers and in the federal role in education research.[4]

Second, the Carl D. Perkins Vocational and Technical Education Act will be up for reauthorization in 2003. A key backdrop to this is the fact that federal priorities and strategic goals for education have increasingly focused on accountability and evidenced-based decision-making as key factors that guide spending priorities. Another important theme in the current administration's effort to shape the federal role in education lies in expanding the flexibility states and localities have in using federal funding. Each of these principles is evident in the reauthorization of the Elementary and Secondary Education Act (No Child Left Behind Act of 2001) and in the U.S. Department of Education's Strategic Plan for 2002-2007.[5] It seems likely that Perkins IV (or its replacement, if there is one) will include much more emphasis on directing resources to maximize improvements in student outcomes and using high-quality research to assess the efficacy of these efforts.

In summary, the relatively small and widely disbursed federal investment in vocational and technical education suggests that these investments should be applied strategically and in ways that maximize leverage on policies aimed at excellence (as well as equity) for secondary school students. There are important gaps in the nation's capacity to undertake a sustained, high-quality research agenda in education and to apply the cumulative knowledge to policy and practice. Although most resources and policy decisions reside with state and local education authorities, they are unlikely to fill this gap at their own initiative. A greater federal role in this area would both maximize federal leverage on policy and practice and fill important gaps left in the landscape of education research and policy making. Finally, the national trends and the current administration's strategic goals and priorities appear to provide unique conditions for a shift in the federal role to provide greater support for systematic innovation and rigorous evaluations.

What should systematic innovation and rigorous evaluation mean?

This section of the paper outlines a general framework for testing promising approaches to vocational and technical education and subjecting them to rigorous evaluations of their impacts, implementation, costs, and benefits. With these principles

---

[3] See http://thomas.loc.gov/ and search for H.R. 4875 in the 106th Congress as cited in National Research Council, *Scientific Inquiry in Education*, p. 93.

[4] National Research Council, *Scientific Inquiry in Education*, pp. 91-112.

[5] See http://www.ed.gov/pubs/stratplan2002-07/

as a guideline, the next section of the paper suggests several specific strategies aimed at moving the federal role in that direction.

A. Systematic Innovation

This paper uses three general principles to characterize systematic innovation: 1) clear and measurable goals for the programs and policies being funded and tested; 2) well-specified "theories of change" that highlight pathways between program components or implementation strategies and goals; 3) adequate flexibility to adapt strategies and components to local needs and circumstances while ensuring adherence to the core goals and theories of change. Principles like these are commonly applied to the design and implementation of specific programs or policies. At a more general level, however, they can also be seen as clarifying federal priorities for investments in vocational and technical education. For example, given that the integration of academic and vocational education is a central goal of the federal investment in vocational and technical education, it would be important not only to clarify what this should mean but also specify how it should be measured. Then, demonstration guidelines, grant applications, or program designs should clearly articulate the chain of events that is expected to advance the goals.

Prioritizing Goals. In specifying program or policy goals in the interest of promoting systematic innovation, it is often useful to consider the question: What would success look like? Or what would need to change or improve if the program were to be judged a success? Put another way, if certain student or program outcomes did not change or improve, the program would clearly be judged a failure. This should include longer-term outcomes such as high school graduation, college enrollment, and labor market entry; short-term outcomes such as test scores and attendance; and mediating factors such as interpersonal supports and classroom instructional practice. It is also important that goals and outcomes be framed as measurable indicators of program benefits or costs.

Specified Program Components and Implementation Strategies. Programs that fall under the umbrella of vocational technical and education are broad, multidimensional, and imbedded within complex layers of school organization and curricula. The absence of well-specified and clearly defined strategies for reaching program goals makes both program design and evaluation difficult. Researchers like Carol Weiss have pointed out, however, that, even when the design in not clearly specified, the policy initiative or program includes implicit "theories of change" that are guiding the program designers, the funders, and, ultimately, the implementers.[6] In an effort to provide better guidance for practitioners in the field and for those who will

---

[6] Carol Hirschorn Weiss, "Nothing as Practical as Good Theory: Exploring Theory-Based Evaluation for Comprehensive Community Initiatives for Children and Families," in *New Approaches to Evaluating Community Initiatives: Concepts, Methods, and Contexts, edited by* James P. Connell, Anne C. Kubisch, Lisbeth B. Schorr, and Carol H. Weiss (Washington, DC: The Aspen Institute, 1995).

evaluate the initiatives, it would be particularly useful for the funding agencies to invest in making its theories of change more explicit and incorporating them into funding guidelines.

Allowing Adequate Flexibility for Adaptation. If education research has offered any consistent lesson in the last one hundred years, it is that policies and programs need to be adapted to local needs and circumstances. In fact, many practitioners, policymakers, and researchers agree that there needs to be a process of mutual adaptation in which the targets of an intervention need to change in order to accommodate its components, while the components themselves need to be modified to accommodate local circumstances. The central challenge lies in allowing adequate flexibility while preserving the basic nature of the innovation being tested. This can be done by not compromising on the central goals of the intervention and ensuring that decisions implementation and adaptation be guided by a common well-specified theory of change.

B. Rigorous Evaluation

For the purposes of this paper, the term "rigorous evaluation" is intended to reflect a set of guiding principles for scientific inquiry rather than a list of research methods. In keeping with general principles that underlie scientific inquiry, the idea of rigorous evaluation should include systematic attempts to identify and control for potential biases that may affect the validity of inferences made about relationships among measured constructs. It should also involve active dissemination of findings, candid descriptions of methods, assessments of the sensitivity of the findings to assumptions underlying those methods, and peer review.

More specifically, the following elements might constitute a working definition of rigorous evaluation in the context of efforts to conduct systematic tests of promising interventions in vocational and technical education: 1) internally valid estimates of the impact of the programs and policies under study; 2) documentation of program implementation and analyses of pathways contributing to goals and outcomes; and 3) measures of net costs and assessment of the relationship between costs, impacts, and implementation. In addition, evaluations of new initiatives should be both summative, answering questions about what works, and formative, answering questions about how and why interventions work or do not work and under what circumstances. In this way, program and research designs should be mutually reinforcing.

Measuring Impacts. At the core of rigorous evaluation should be a reliable, internally valid assessment of impacts. Here, it is critical to differentiate between outcomes and impacts. *Outcomes* refer to measures of student, teacher, or organizational behavior or functioning. For students, this might include measures of engagement, achievement, educational attainment, or attitudes. For schools, this might include aggregate measures of teacher turnover, student achievement, or graduation rates. A key challenge, noted in the discussion of systematic innovation above, lies in

determining which outcomes might serve as the best indicators of particular program or policy goals.

Focusing exclusively on outcomes, however, presents the evaluation effort with risks that can render the findings confusing, at best, and misleading, at worst. Consider, for example, Manpower Demonstration Research Corporation's (MDRC's) evaluation of high school Career Academies whose primary goals include preventing students from dropping out of high school and helping them make successful transitions to postsecondary education and the labor market.[7] Career Academy programs select students who are already engaged in school and have high aspirations to go on to a two-year or four-year college. This resulted in a high percentage of participating students graduating and going on to college, but these were also students who were highly likely to do so regardless of whether they were in an Academy. By contrast, the Career Academy programs also included poorly motivated students with no particular orientation toward postsecondary education. In absolute terms, a relatively low proportion of these students graduated from high school and went on to college. Yet, this represented a somewhat higher percentage compared to a similar group of students who did not enroll in the Academies.

Impacts, therefore, refer to the effect that an intervention, program or policy has on an outcome, over and above what would have occurred without the program or policy. Addressing questions about impacts requires that outcomes for students (or teachers or schools) exposed to a given program be compared with outcomes for truly similar students (or teachers or schools) not exposed to it. Only this comparison can shed light on the extent to which the program really made a difference. The central problem in impact research lies in identifying a truly comparable "counterfactual" or control group that adequately represents a condition where the program is, or was not available to students, teachers or schools who were not systematically different from those who were exposed to the program.

Random assignment offers the most reliable approach to identifying a truly comparable counterfactual or control group. Others have argued passionately and eloquently for wider, conscientious use of random assignment designs in education.[8] In fact, field experiments using random assignment are used much more frequently in other policy arenas like employment and training and welfare-to-work than in education.

---

[7] James J. Kemple, *Career Academies: Impacts on Students' Initial Transitions to Post-Secondary Education and Employment* (New York: Manpower Demonstration Research Corporation, 2001); James J. Kemple and Jason C. Snipes, *Career Academies: Impacts on Students' Engagement and Performance in High School* (New York: Manpower Demonstration Research Corporation, 2000).

[8] See, for example, Donald T. Campbell and J. C. Stanley, *Experimental and quasi-experimental designs for research* (Chicago: Rand-McNally, 1963); Robert F. Boruch, *Randomized Experiments for Planning and Evaluation* (Thousand Oaks, CA: Sage Publications, 1997); and Thomas Cook, "Considering the Arguments Against Random Assignment: An Analysis of the Intellectual Culture Surrounding Evaluation in American Schools of Education" (paper presented at the Harvard Faculty Seminar on Experiments in Education, Cambridge, MA, 1999).

This paper concurs with the call for greater use of random assignment in education evaluation research but will leave it to the reader to draw on the sources listed below for discussion of its methodological advantages and disadvantages. For the purposes of this paper, it might be useful to offer a few notes on conditions that are typically necessary for its successful application and implementation.

Over 25 years, MDRC has conducted more than 30 major random assignment evaluations in more than 200 locations, involving close to 300,000 individuals. This experience has yielded a number of lessons about how to successfully mount random assignment evaluations (of individuals and clusters of individuals). It also points to the vulnerability of this type of research and the need for care and sensitivity in its use. Following are several key challenges that must be addressed in mounting a successful random assignment evaluation.[9]

- **Addressing Questions Primarily About Net Impact.** The first criterion for deciding the appropriateness of using random assignment is whether the central research question is about the extent to which the intervention under study causes changes in the outcomes it was designed to affect. Questions about feasibility, replicability, or explanations of successful or failed implementation are not typically good candidates for random assignment.

- **Meeting Ethical and Legal Standards and Requirements.** Most notably, this means ensuring that those being randomly assigned (students, teachers, classrooms, schools, etc.) are not denied services to which they would otherwise be entitled. In some instances, however, there are not enough resources to serve everyone who would be interested and qualified to participate in a given program. Random assignment may be a fair and rational way to allocate scarce resources. Other initiatives may involve changes in requirements for participation in existing programs, and random assignment may be a fair and rational way to phase in these requirements before they are applied to all students.

- **Building a Consensus Among Key Stakeholders that Random Assignment is the Best Available Alternative and that the Value of the Study Outweighs the Added Burden that May Be Expected of them.** There is no replacement for intellectual and political consensus-building in mounting a successful study. Concrete incentives may also be needed in cases where participation is purely voluntary and may involve extra costs for data collection or effort to sustain involvement in the treatment or control group.

---

[9] These lessons are captured in greater depth in Judith M. Gueron, "The Politics of Random Assignment: Implementing Studies and Impacting Policy" (paper presented at the American Educational Research Association Annual Meeting, 2001, available from the author at Manpower Demonstration Research Corporation, New York).

- **The Intervention has (or will have) Observable Features that Distinguish it from Alternative Education Services and Programs (including the status quo).** To the extent that an innovative program cannot be distinguished from existing programs or services (or its distinguishing features have not been well specified) there are probably more pressing research questions about feasibility and program design that should be addressed first. In many cases, however, the evolving magnitude and quality of differences between the program and its alternatives are themselves useful targets for investigation in a random assignment study. It is also important to have some assurance that the subjects in a random assignment study are treated in ways that are consistent with their status in the study. This will help maximize the expected contrast between the treatment and control statuses.

- **Collecting Data on Outcomes that Reflect Intended Goals of the Intervention and Assuring that Data Collection does not Inject Bias Into the Comparability of the Treatment and Control Groups.** Often it is assumed that the central goals of random assignment have been met once the coin has been tossed and the treatment and control groups have been established. As with nonexperimental designs, a key source of bias can arise when certain data are available for one group and not for the other or when some constructs are measured more reliably for one group than for another. Random assignment has the distinct advantage of eliminating such bias at the point where the treatment and control groups are initially determined. This can be undone, if later data collection and measurement is conducted differentially for each group.

In addition to providing a sense for the ingredients of success, the challenges listed above also help illustrate conditions under which random assignment may not be appropriate or feasible. Random assignment should be seen as a tool for answering questions about the impact (not just the outcomes) of reasonably well-defined interventions. It is most useful in cases where alternative methods for selecting people or schools into the intervention would confound the attribution of effects. In other cases, however, random assignment may not be necessary when other methods can identify a counterfactual that yields a "reliable enough" answer to questions about the impact of a program or policy intervention. For example, it may be possible to eliminate alternative explanations for changes in outcomes that are coincident with the implementation of a new program. Here, a school's past record of student outcomes may serve as the counterfactual for comparison to outcomes subsequent to the start of a new or upgraded vocational education program.[10] While alternative designs like interrupted times series analyses can be less intrusive than random assignment, their rigor and reliability are a function of many of the same challenges listed above.

---

[10] For a discussion of the statistical assumptions and properties of interrupted time series designs, see Howard S. Bloom, "Measuring the Impacts of Whole-School Reforms: Methodological Lessons from an Evaluation of Accelerated Schools" (New York: Manpower Demonstration Research Corporation, 2001).

Finally, it is also important to note that, even if all the challenges to random assignment are addressed and the design can be sustained, the value of the findings still hinge, in large part, on the capacity of systematic process and implementation research to illuminate the likely underlying sources of difference between the treatment and control conditions (getting inside the "black box").

Implementation and Process Research. In conducting rigorous evaluations of promising initiatives, it is not enough to know only *whether* it had an impact on student, teacher, or school outcomes. It is also important to learn as much as possible about *why (or why not), how, and what might be done next.* In short, impact findings are likely to be most useful in the context of information about: 1) whether the program, both in theory and as implemented, was adequately equipped to make a difference for the target population (schools, classrooms, students, teachers); 2) how the program changed (or did not change) local conditions and the educational experiences of the target population along the expected pathways leading to its ultimate goals; and 3) whether, and if so, how the program should be adapted to a wider range of conditions. While exploration of these issues may have value independent of their role in an impact evaluation, their utility can be greatly enhanced when impact and implementation analyses are integrated. Following is a brief overview of three key interrelated dimensions of implementation research that is conducted in the context of impact studies.[11]

- **Assessing the Theoretical and Empirical Basis for the Program's Capacity to Produce an Impact.** As with systematic program designs, rigorous implementation research should be built on making the theory of change underlying the program as explicit as possible. Promising vocational and technical education programs may be complex, but it is critical to clarify both how it is different from pre-existing programs or conditions and how that difference is supposed to accomplish the program's key educational and employment goals. Ideally, program and evaluation design can occur simultaneously to ensure that the program has the maximum potential to affect change and the research is aimed in the right direction with the right data collection and analysis tools. Empirically, assessing a program's capacity to produce an impact amounts to investigating the extent to which it received a "fair test" in the field. Typically, this includes documenting the core components and services encompassed by the intervention and tracing the pathways through which they change the educational environment and experiences of the target population. Then, the implementation research focuses on determining whether the program reached the intended target population and whether it provided them with an adequate "dosage" of

[11] The material in this section of the paper is based on ideas discussed in Key E. Sherwood and Fred Doolittle, "What's Behind the Impacts: Doing Implementation Research in the Context of Program Impact Studies" (unpublished paper available from the authors at Manpower Demonstration Research Corporation).

intended treatment. Finally, impact-driven implementation research is distinct from process research conducted outside this context because it must aim to assess the extent to which the program, as implemented, reflects services and experiences that are truly different from those experienced by those in the control group. In many ways, this is even more important than documenting the program's fidelity to its ideal model.

- **Identifying Sources of Impacts and Reasons for a Lack of Impacts.** Impact evaluations are sometimes characterized as "black box" evaluations where people or organizations are randomly assigned as equal entities at one end and the differences that emerge at the other end represent changes that were mysteriously created inside. Getting inside the black box is more challenging methodologically, and involves less certainty, than impact analysis, particularly in education research where the black box (schools, classrooms, and so on) is especially messy. Yet, this enterprise is crucial to maximizing the policy and practical relevance of the impact findings. High-quality impact-driven implementation research should be driven by theory, past evaluations of related interventions, and by observation and interaction with the target population and those delivering services. Again, theory is an essential starting point in directing implementation researcher's attention to the most likely (or, at least, the intended) source of impacts and the most likely factors that would mitigate a positive impact. Prior research serves a similar purpose while helping to place the current evaluation and its findings in a broader context. Finally, there is no replacement for drawing on the authentic and immediate experiences and perspectives of the individuals who are engaged in the program being studied. In the context of an impact evaluation, the experiences and perspectives of those in the control group are equally important.

- **Investigating Factors that Can Guide Further Adaptation or Replication.** A central goal of an impact evaluation is usually to produce a "bottom line" on whether policymakers or practitioners should continue to support a program or expand its use. Often, however, more nuanced responses are required to this question, regardless of whether the impact findings are positive or negative. For example, a vocational education program that shows strong positive impacts on standardized test scores may not be a good candidate for replication if the results were produced primarily by ignoring other goals and allocating a disproportionate amount of classroom and non-classroom time to test preparation. By the same token, another vocational education program that had little or no impact on postsecondary education and employment outcomes should not necessarily be discarded if only a small proportion of the target population received the intended treatment for reasons that could be (or were) addressed with alternative implementation strategies. Thus, the job of implementation research in the context of impact evaluations should be to not only document whether it might be worthwhile to expand the use of the program under study, but to provide

empirical evidence about how this might be done and how the program would need to be adapted or modified as it is introduced into new localities and circumstances.

Finally, it is important to recognize that implementation research in the context of an impact evaluation differs from a typical, stand-alone process or implementation study primarily in the degree to which it is a deductive or hypothesis-testing driven enterprise rather than an inductive or primarily descriptive undertaking. Like the impact evaluation of which it is a part, impact-driven implementation and process research requires extensive forethought regarding the hypotheses it aims to test and the data that should be brought to bear on those hypotheses. Implementation researchers should, of course, be able to surface unanticipated issues and findings in the course of fieldwork or data analysis. In general, however, the goals of impact evaluations are generally less well served by implementation research techniques whose central goals are to develop or shape the central research questions and hypotheses during fieldwork and analysis. At the same time, impact-drive implementation research should use multiple methods (qualitative and quantitative) and should be prepared to rely on rich description and case studies to help illustrate what is behind the impact findings.

Cost, Cost-Benefit, and Cost-Effectiveness Analyses. Cost-benefit analyses have been used fairly extensively to guide the more efficient use of resources in many areas of public policy, including investments in human capital that are less directly in the purview of federal support for vocational and technical education, such as employment and training and welfare-to-work policy. In education, however, there are few examples of studies that include systematic measures of an intervention's cost (and, for evaluation purposes, its *net* cost, over and above viable alternatives, including the status quo) and an assessment of the balance between net costs and net benefits or impacts.[12] It might be argued that this line of inquiry is not useful because it is difficult to attach monetary values to many important benefits of vocational and technical education. Even if this were true, the field should make greater use of studies that weigh net costs against non-monetized impacts to assess an intervention's cost-effectiveness. Typically, however, costs are accounted for primarily through audits of authorized expenditures, which shed little light on net costs or the benefits that accrued from that investment. Over time this has left the field with scant evidence of likely tradeoffs of investing in alternative vocational and technical education strategies.

For guidance on conducting systematic analysis of programs costs and assessments of the relationship between costs and program benefits and effectiveness, rigorous evaluations of vocational and technical education will need to turn to other policy domains – most notably evaluations of welfare-to-work and employment and

[12] Anthony E. Borardman, David H. Greenberg, Aidan R. Vining, and David L. Weimer, *Cost-Benefit Analysis: Concepts and Practice*, Upper Saddle River, NJ: Prentice Hall, 1996, pp. 445-472.

training programs.[13] As with education, programs that have been evaluated in these fields involved layered, multistream funding arrangements; complex, interactive services; and difficult challenges for measuring and monetizing costs and benefits. The reader is referred to the sources listed below for an overview of the cornerstones of systematic cost analysis leading to assessments of the relationship between net costs and benefits (monetized impacts) or effectiveness (non-monetized impacts).

Toward Strategies for Shifting the Federal Role to Focus More on Systematic Innovation and Rigorous Evaluation

So far, this paper has attempted to present a rationale for a federal role in vocational and technical education that would be more focused on systematic innovation and rigorous evaluation and to highlight some core principles that might underlie such a role. This section of the paper outlines three interdependent strategies that may serve to shift federal investments in this direction. The ideas presented here reflect a work in progress and require more depth in order to confront the political and level realities of the current, and likely future, federal role in vocational and technical education.

Making Larger Investments in Fewer, More Targeted, and Well-Specified Initiatives. Current legislation and ED policies already focus on supporting innovation and rigorous evaluation. For example, Perkins III explicitly points to reform, innovation, and continuous improvement as central goals of the legislation and its implementation. Yet, approximately 85 percent of Perkins funding is distributed to states based on formulas and regulations that make it difficult to leverage reform, innovation, and continuous improvement. This places a high premium on the utility of remaining resources that are invested in NAVE and the national centers. Pending significant changes in legislation, the Office of Vocational and Adult Education's (OVAE) demonstration authority provides the most promising route to these goals. Given the limited funding available, the federal government is likely to gain greater leverage on policy and practice by making larger investments in a few large-scale initiatives that would yield credible results regarding the impact of federal investments in vocational education programs on students' postsecondary education and labor market outcomes.

For example, OVAE might support differential impact studies of various configurations of vocational education resources and programs. This might involve the following types of experiments and demonstrations:

---

[13] See David H. Greenberg and Ute Appenzeller, *Cost Analysis Step by Step: A How-to Guide for Planners and Providers of Welfare-to-Work and Other Employment and Training Programs* (New York: Manpower Demonstration Research Corporation, 1999); in addition to Boardman, Greenberg, Vining, and Weimer, *Cost-Benefit Analysis: Concepts and Practice.*

- A study comparing vocational education alone to vocational education plus specific enhancements such as high-quality work-based learning experiences. This would enable us to estimate the benefit of exposing students to high-quality work environments as part of their secondary school experience.

- A study comparing standard vocational education programs to employer-driven vocational educational programs, such as those that include active occupationally specific advisory councils, teacher summer internships with employers, student work-based learning opportunities, sectoral strategies in high-growth areas, and employer-led work readiness classes in schools.

- A test comparing standard vocational education to enhanced professional development for academic and vocational teachers centered on developing integrated curricula and instructional practices to best teach the curricula.

- A study comparing alternative versions of career-related school-restructuring programs like Career Academies, career clusters, or Tech Prep to "standard" vocational education typically comprised of sequences of occupation-specific or technical content courses.

- OVAE might also consider partnering with the Department of Labor's Employment and Training Administration to conduct systematic tests of programs that blend resources for the Workforce Investment Act and Perkins III.

Some elements of Perkins III reflect several concrete steps toward placing a stronger emphasis on program innovation, coupled with rigorous evaluation, in an effort to increase the body of knowledge about which programs work and why. For example, it relaxed a number of funding restrictions and mandated an "independent evaluation and assessment of vocational and technical education programs under this Act." Notably, in response to the growing national interest in Tech Prep programs, Perkins III authorized, and OVAE later implemented, a national demonstration of Tech Prep programs in which high school programs would be housed on community college campuses. OVAE solicited proposals for awards to fund programs (including development and evaluation) over three years. The Tech Prep Demonstration Program (TPDP) provides a useful prototype for focusing federal resources on innovative vocational and technical education strategies and submitting them to rigorous evaluation.

Coordinating Program Development and Research Designs. Targeting resources to particular tests of promising strategies and rigorous evaluations of their impacts, implementation, and costs would be a useful first step in maximizing opportunities to learn. A second step would involve carefully coordinating the design of the interventions to be tested with their evaluation designs. A critical challenge would involve balancing prescriptiveness, to ensure sufficient comparability across sites,

against flexibility, to allow adaptation to local needs and circumstances. Following are several guiding principles that might be used in constructing a test of promising vocational and technical education strategies:

- Set priorities regarding learning goals including short-term and long-term outcomes, target populations, and expected alternatives to which the intervention should be compared (including the status quo).

- Make the theory of change underlying the intervention explicit to guide implementation strategies, clarify hypotheses for the evaluation, and identify critical targets for evaluation measurement and data collection.

- Specify research methodologies that are consistent with the learning goals and the underlying theory of change.

- Require coordination among program designers, implementers, and evaluators as a condition of the grant or contract.

At a general level, many federal requests for proposals include some elements of these principles. Typically, however, many of these issues are left for the applicants to specify. Lack of clarity about goals, the underlying logic of the intervention, and coordination of program and evaluation design will severely limit learning opportunities and longer-term benefits to policy. The TPDP request for proposals may provide a example where some of the design principles outlined above were specified while others were left open and could have been made more explicit or prescriptive.

On the one hand, TPDP built on the fairly specific criteria that Perkins III set for Tech Prep program elements and goals by requiring adherence to particular components of an eligible consortium and requiring the location of the secondary education program on a community college campus. TPDP also required third-party evaluations of both implementation and impacts using "rigorous, scientifically accepted methods." It also placed a high priority on high school graduation and transitions to postsecondary education and employment as the key goals of the programs and as outcomes to be measured in the evaluation.

On the other hand, TPDP allowed each program to design and conduct (or contract) an independent evaluation. As an alternative strategy, it might have called for conducting a single national centralized evaluation of the winning projects. This strategy may have offered several advantages: 1) standardization of the research design and centralization of data collection and analysis, resulting in a more rigorous evaluation and possible monetary savings; 2) allowance for research designs not feasible in a single site, including the random assignment of sites to a project or control group; 3) allowance for projects to focus resources on development and implementation of high-quality programs. In short, a single national evaluation conducted by an

independent evaluator on behalf of OVAE has the potential to be much more rigorous, informative, and powerful than ten separate evaluations conducted, potentially, by ten different evaluators with ten different designs, ten sets of data, ten separate analyses, and potentially ten separate conclusions. The TPDP request for proposals also offered little guidance about the coordination of research and program designs. By clarifying the need and some strategies for greater coordination, the operating consortia and program implementers would be assured of having input on what the evaluation should measure and how. This would also provide the evaluators with more access to the development of the intervention and enable them to integrate evaluation design features and data collection procedures into normal program operating procedures.

Incentives to States, Local Education Authorities, and Schools. Having set some priorities for a more targeted federal investment vocational and technical education and having clarified the rules under which this investment would proceed, the next key question is how to get the key actors to play. While some key stakeholders at the state, local, and school level are likely to be motivated by the opportunity to contribute to the field and general knowledge-building, others may need incentives that that hit closer to home. Such incentives are likely to come in the form of larger funding amounts, alternative rules about how the funding can be used, or a combination of the two.

On the one hand, assuming that larger overall allocations for vocational and technical education would not be forthcoming, additional funding may be possible if the shift in federal focus involved redirecting existing resources into larger, but fewer and more targeted demonstrations and evaluations. Participating districts or schools could qualify for a 10 percent bonus to participate in a serious research effort, part of which would be used to compensate for research-related costs such as data collection, accommodating field research, and participating in dissemination and consensus-building activities. A critical task here would involve working closely with grant or contract recipients to clarify the prescribed use of resources and to identify areas of flexibility for adapting their use to local needs and circumstances.

On the other hand, states and districts could be granted waivers for more flexible use of vocational education resources. For example, schools and districts might be allowed to waive or modify their performance standards as an incentive to target recourses for students at particularly high risk of school failure. Some programs or schools may not otherwise place a high priority on serving such students, particularly if their performance would be penalized by their higher propensity for dropping out, performing poorly on standardized tests, or failing to make a successful transition to postsecondary education and the labor market. Such waivers would be granted, however, only if the schools and programs agreed to evaluate the rules being waived (this would be akin to the Section 1115 research requirements of the Social Security Act prior to the passage of the Personal Responsibility and Work Opportunity Reconciliation Act in 1996).

Conclusion: Beyond Vocational Technical Education

As noted at the outset, a federal commitment to systematic innovation and rigorous evaluation should extend more broadly beyond vocational and technical education to elementary and secondary education. The recently reauthorized Elementary and Secondary Education Act places an unprecedented emphasis on accountability and on identifying educational strategies that are "proven" to work. At the same time, increased flexibility offered to states, local education authorities, and schools could provide a unique opportunity to test promising ideas in the pursuit of both accountability and excellence. A critical challenge resides in ensuring that the new federal government investments in elementary and secondary education yield more reliable evidence about what works and why than has previously been the case.

Useful steps have been taken in this direction, but more are needed. Over the past several years, for example, the U.S. Department of Education has called for the use of "research-based" approaches to school improvement. Yet, the standards of evidence for this research base are not clearly defined, and there has been little systematic effort to align research and program designs. For example, initiatives like the Office of Educational Research and Improvement's (OERI's) Comprehensive School Reform scaling-up and capacity-building contracts and grants call for third-party evaluations but provide little guidance on how to ensure that the evaluations can offer findings and lessons from across the demonstration sites and program models. Subsequent support for the six cross-cutting evaluations were aimed at trying to fill this gap, but several of these do not appear to be well coordinated with the model developers, and the research designs are quite varied even though there is a great deal of overlap in the research agendas.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

	This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.
X	This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").